12-2006

Measuring Qualities of Articles Contributed by Online Communities

Ee Peng LIM
Singapore Management University, eplim@smu.edu.sg

Ba-Quy VUONG
Nanyang Technological University

Hady W. LAUW
Singapore Management University, hadywlauw@smu.edu.sg

Aixin SUN
Nanyang Technological University

DOI: https://doi.org/10.1109/WI.2006.115

Measuring Qualities of Articles Contributed by Online Communities

Ee-Peng Lim, Ba-Quy Vuong, Hady Wirawan Lauw, Aixin Sun
Nanyang Technological University
Nanyang Avenue, Singapore

Abstract

Using open source Web editing software (e.g., wiki), online community users can now easily edit, review and publish articles collaboratively. While much useful knowledge can be derived from these articles, content users and critics are often concerned about their qualities. In this paper, we develop two models, namely basic model *and* peer review model*, for measuring the qualities of these articles and the authorities of their contributors. We represent collaboratively edited articles and their contributors in a bipartite graph. While the basic model measures an article's quality using both the authorities of contributors and the amount of contribution from each contributor, the peer review model extends the former by considering the review aspect of article content. We present results of experiments conducted on some Wikipedia pages and their contributors. Our result show that the two models can effectively determine the articles' qualities and contributors' authorities using the collaborative nature of online communities.*

1 Introduction

1.1 Motivation

Web users today form different online communities among themselves for information sharing and collaboration. One most notable trend is the use of wiki software to operate websites (known as wikis) for users to collaboratively edit content easily without having to know HTML and to standardize the look and feel of these websites[10]. Among the large and more successful wiki sites is Wikipedia[9], the online encyclopedia which covers 3.7 million articles (both English and non-English), 45,000 registered users, and 200 languages. As the 37th most visited website in WWW, Wikipedia represents a valuable resource for learning.

While wiki sites are growing fast in number and size, there are also some serious concerns over the quality of content found at these sites. Many wiki sites (including Wikipedia) allow anyone to edit and contribute content, often without requiring the contributors to register themselves. These contributors are not paid and their expertise may not be verified. It is therefore not clear how the collaborative editing process used in wikis can ensure accuracy and authenticity in their article content.

Wiki advocates often counter-argue that having more pairs of eyes and ease of making corrections help to weed out errors in wiki content. A special recent investigation by Nature also suggested that Wikipedia can match Britannica (an established encyclopedia) in terms of accuracy for its science articles[4]. These observations however do not address a more pertinent problem, that is, the measurement of qualities of wiki articles.

Distinguishing between good and bad quality articles is not a simple task to human users, let alone computer programs. The difficulties can be attributed to several reasons, namely:

- *Large number of articles for quality judgement*: Ironically, the larger the wiki site, the harder is to determine the quality of each article by comparing with other articles from the same site.
- *Diverse content among articles*: Wide range of topics can be covered by the articles. It is extremely difficult to perform content analysis on the article to determine their qualities without human judgements and high quality benchmark collection for each topic.
- *Unknown contributors*: As mentioned earlier, the expertise and experience of contributors are usually not explicitly captured by the collaborative software. Without knowing this, it is difficult to determine the quality of articles created by them.
- *Abuse*: Wiki sites with open access can easily be targets of abuse when contributors can intentionally cre-

ate articles of specific patterns to circumvent quality checking. In this case, a human expert may be able to detect such instances but designing a software to detect them will be a challenge.

Our research therefore aims to automate the measurement of qualities as much as possible, without interpreting the article content. Instead, we will only draw clues from collaboration and edit history. We represent a collection of articles contributed by online users by a bipartite graph as shown in Figure 1. The graph consists of a set of contributors (u_i's) and a set of articles (r_j's). We are only interested in the quality of the *latest* versions of articles. The directed edges from contributors to articles represent contributions. Each edge is assigned a value c_{ij} representing the amount of contribution. The c_{ij} value can be measured in terms of number of words. Again, the contribution refers to that for the latest version of article. Those portions of content removed by contributors prior to the current version are therefore not considered.

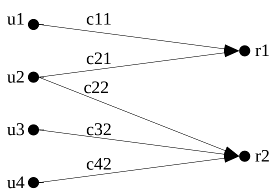

Figure 1. Example Scenario (with c_{ij} values)

In the absence of manual judgement, one can try to measure article quality by the number of words it contains. The longer the article is, the more quality it is expected to carry. This approach, known as the **naive model**, can be expressed by Equation 1.

$$Q_j = \sum_i (c_{ij}) \tag{1}$$

The naive model clearly has major shortcomings. It does not consider the quality of the contribution and can be easily abused. For instance, one can create very long articles to get them assigned with high qualities.

1.2 Objectives and Contributions

The objectives of this research is to investigate novel models for measuring qualities of collaboratively created articles. Our idea of designing models for calibrating the quality of articles is based upon the following principle which we call the *mutual reinforcement principle*:

- *Quality: An article has high quality if it is contributed by high authority authors.*
- *Authority: A contributor has high authority if s/he contributes high quality articles.*

The mutual reinforcement principle also suggests that as we measure article qualities, the contributor authorities will also have to be determined.

The naive model in Equation 1 is clearly not designed based on the mutual reinforcement principle, although it can also define the authority of a contributor by the total amount of contribution they make to articles (see Equation 2). Since the qualities of contributed articles are not considered, it is possible that some contributor acquires much authority by creating very long articles or many articles.

$$A_i = \sum_j (c_{ij}) \tag{2}$$

In the following, we summarize our contributions:

- We develop two models for measuring article qualities and contributor authorities based on the mutual reinforcement principle and they are known as the *basic model* and *peer review model*.
- We describe the implementation of the two proposed models and adopt an iterative computation approach for these models.
- We evaluate the two proposed models on some Wikipedia articles and contributors. The different quality (authority) rankings produced by the naive, basic and peer review models are compared and some interesting examples are given.

Usually, a community created article evolves over time as it accumulates content from different contributors. This implies that its quality also varies with time. For practical reasons, we are only interested in measuring the quality of articles in their present or latest versions since they are often the ones perused by readers, not the historical versions.

1.3 Paper Outline

The rest of the paper is organized as follows. Section 2 describes some related research. We present our proposed basic model and peer review model in Section 3. The implementation issues are examined in Section 4. We then describe the experiments conducted to compare the models and present the results in Section 5. Finally, we conclude the paper in Section 6.

2 Related Work

Quality measurement for web content is a new challenging research topic. In general, there are several different approaches to treat quality.

One can hire human experts to select quality web pages for indexing as in the case of web directories (e.g., Yahoo! directory[1]). This approach clearly is not scalable and ranking pages by quality is also an extremely difficult problem for human experts.

For web pages, the two link analysis models, page rank[8] and HITS[6], exploit the recommendation semantics of inter-page links to derive page rank, hub and authority values of web pages as some forms of quality measurement. Page rank and HITS have been applied to web search by combining with some content-based search techniques[3], and to web navigation[8] and web crawling that focus on high quality pages[7]. The relationships between page rank, authority and quality have been studied by Amento, Terveen and Hill[1]. The work concluded that the link analysis models did well in measuring quality. To detect web spam pages which muster excessive page ranks by artificially created links, Gyongyi, Garcia-Molina and Pedersen developed the TrustRank algorithm to identify good pages from bad ones using a small set of seed pages[5].

The above models however only apply to a network of web pages. In the case of wiki sites, the links among wiki articles do not always carry recommendation semantics. More often, they merely provide the links to some definitions. Nevertheless, Bellomi and Bonato had applied PageRank and HITS on Wikipedia articles and recognized some organization structure and cultural biases in the linkage[2]. The above model however does not recognize that there are actually two types of entities in the network, i.e., articles and contributors, and their interactions in quality measurement.

While our proposed basic model borrows some ideas from HITS, we further extend it to the peer review model which draws information from the collaboration among online community users.

3 Proposed Models

3.1 Basic Model

We now describe a basic model that considers an article as the aggregated efforts of multiple contributors playing the authoring role. The mutual reinforcement principle is realised by examining not only the number of articles, but also the amount of content (c_{ij}'s) authored by a contributor. In the basic model, the article quality and contributor authority are defined in Equations 3 and 4 respectively.

$$Q_j = \underset{i}{\mathcal{A}gg}\,(c_{ij} \cdot A_i) \tag{3}$$

$$A_i = \underset{j}{\mathcal{A}gg}\,(c_{ij} \cdot Q_j) \tag{4}$$

[1]http://dir.yahoo.com/.

In the above equations, c_{ij} represents the amount of content measured by a function of the *number of words* in article r_j authored by contributor u_i. Equation 3 says that article quality is an aggregation of the authorities of contributors multiplied with the number of words from the contributors. The aggregate function, $\mathcal{A}gg$, represents a class of functions to combine the contributing authorities (or qualities). While $\sum$ is an obvious choice, there can be other alternatives that may serve well. Furthermore, the two $\mathcal{A}gg$'s used in the model (and also in other models) are not necessarily the same. We will discuss the choices of $\mathcal{A}gg$ in greater detail in Section 4.

3.2 Peer Review Model

An article may undergo a series of changes by different contributors. For each change to be made by a contributor, he or she will have to first review the prior content of the article and decide which parts to add or remove. In other words, the text content in the article that survives the change is more or less approved by the current contributor. This therefore leads us to design the *peer review model* for defining quality and authority.

$$q_k = \mathcal{A}gg(A_{i_k}, \bigcup_{u_i\ reviews\ w_k} \{A_i\}) \tag{5}$$

where u_{i_k} denotes the author of w_k and A_{i_k} represents his/her authority.

$$Q_j = \underset{w_k \in r_j}{\mathcal{A}gg}\ q_k \tag{6}$$

$$A_i = \underset{j}{\mathcal{A}gg}\,(c_{ij} \cdot Q_j) \tag{7}$$

The peer review model represents an article by a *bag of words*. In Equation 5, we define the notion of *word quality* denoted by q_k. Each word w_k in an article is first created by some author (or contributor) u_{i_k} who has an authority of A_{i_k}. w_k is subsequently "reviewed" by a series of other contributors u_i's of the article containing w_k. Under the peer review model, we assume that this kind of review is carried out as the subsequent contributors add other parts of the article. This assumption is quite reasonable as contributors usually have to read through articles before making changes. The aggregate function $\mathcal{A}gg$ derives an aggregated authority for w_k out from the authority (A_{i_k}) of the contributor who authored w_k and the authorities (A_i's) of other contributors who reviewed w_k. Ideally, we would want the word authority to reflect the quality of w_k. Since the roles of authors and reviewers are different, the aggregate function takes them as two different input parameters. A detailed discussion of this aggregate function is given in Section 4.

Equation 6 is an aggregation of word qualities for those words contained in an article r_j. This gives the overall quality for r_j. Here, $\sum$ is a good candidate aggregate function but others are possible. Note that Equation 7 remains the same as Equation 4 as the peer review model focuses only on capturing the review information embedded in the edit histories of collaboratively authored articles.

Interestingly, if all articles are not "reviewed" at all in cases when the subsequent contributors do not need to read the prior content or the aggregation function $\mathcal{A}gg$ ignores the reviewers' authorities completely, the second component of Equation 5 ($\cup_{u_i\ reviews\ w_k} \{A_l\}$) will disappear and the peer review model will reduce to the basic model.

4 Implementation Issues

4.1 Aggregation of Contributor Authorities

Quality in Basic Model

In the basic model, article quality is an aggregation of products between contributor authorities A_i and amount of contributions c_{ij}. A simple choice for the function $\mathcal{A}gg$ in Equation 3 is $\sum$ which sums up the authorities of all contributors treating everyone the same.

$\mathcal{A}gg = \sum$, however, may be unfairly exploited when an article is created by large number of low authority contributors who artificially inflat the quality of the article. This may be addressed by using $\mathcal{A}gg = Max$ which uses only the authority of the most authoritative contributor, or $\mathcal{A}gg = \sum_{top\ k}$ which sums the authorities of the k most authoritative contributors.

For simplicity, we will use $\mathcal{A}gg = \sum$ to combine the contributors' authorities for the rest of the paper.

Quality in Peer Review Model

As shown in Equation 5, the peer review model defines word quality as an aggregation of the authorities of word's author (u_{i_k}) and reviewers. This aggregate function $\mathcal{A}gg$ has to deal with authorities of both authors and reviewers. In one extreme, it can be defined to ignore the reviewer component completely reducing the peer review model to the basic model as mentioned earlier.

The following are some possible ways to define the function:

- *Author and reviewers are all important ($\sum$):* By considering the author and reviewers equally important, we can sum their authorities up to represent the word quality.
- *Use of champion(***TOP_AR***):* In this alternative, we choose the maximum authorities among the author and reviewers. That is:

$$\mathcal{A}gg(A_{i_k}, \bigcup_{u_i\ reviews\ w_k} \{A_i\}) = Max(A_{i_k}, \bigcup_{u_i\ reviews\ w_k} \{A_i\}) \quad (8)$$

Other than $\mathcal{A}gg = \sum$, the other alternative aggregate functions unfortunately have some implications to the computation of peer review model as the iterative computation of matrix equations involving these functions may not converge. A detailed study of their convergence is an interesting direction for further research. Hence, we have adopted $\sum$ as the aggregate function for word quality for the rest of this work.

4.2 Aggregation of Article Qualities

In both the basic and peer review models, the computation of a contributor authority requires an aggregation of the qualities of articles weighted by the amount of contribution from him or her. The aggregate function $\mathcal{A}gg$ in Equation 4 can be a simple $\sum$ that adds the weighted qualities concerned together. We can also explore other ways of aggregating the article qualities, say $\mathcal{A}vg$, $\sum_{top\ k}$, etc.. For simplicity, we will use $\sum$ for the rest of the paper.

4.3 Iterative Computation of Proposed Models

The implementation of our two proposed models can be formulated as a matrix computation. The convergence of their iterative computations to a unique solution is assured for different initial quality values. In the computation of article qualities, word qualities and contributor authorities, we apply L_1 normalization on the values. For example, the L_1 normalization of vector U_A refers to adjusting A_i's such that $\sum_{i=1}^{n} |A_i| = 1$. The normalization preserves the relative ratios among vector elements, and will not affect the convergence properties of our proposed models.

5 Experiments

5.1 Data Set

Our data set consists of 77 articles each about a randomly selected country. These articles and their edit histories were downloaded on 19 June 2006. These country names were obtained from a publicly available listing of 243 countries.

All stop words in the articles (e.g., "a", "the", "in", etc.) were removed. Next, we extracted all contributors who contributed to the current versions of these articles (after stop

Table 1. Statistics of Dataset

# articles	77	# contributors	1083
# contributors/article	21.58	# articles/contributor	1.53
# terms/article	2783.1	# terms/contributor	197.88

word removal). There are altogether 1083 contributors. To implement the peer review model, we also derived the author and reviewers of each word in the article according to edit history. The essential statistics of our dataset is shown in Table 1.

According to the Wikipedia measurement done by Voss[9], a third of contributors have only contributed in one article. In our data set, we also found that most of the contributors contributed to only one article as the average number of articles contributed is 1.53. This suggests that majority of contributors will likely get low authority values. On the other hand, there are on average 21.58 contributors contributing to each article showing a significant level of collaboration among contributors.

As the number of words contributed varies very significantly across different contributors, we have chosen to dampen the amount of contribution c_{ij} by applying the log function on the number of words. This additional damping mechanism has impact on the final computed values but it does not affect the convergence properties of our proposed model.

5.2 Overall Results

All our experiments involved a small number (<30) of iterations and the iterative computation converged in few seconds for the dataset. Convergence was considered to have achieved when the delta changes to the quality and authority values were smaller than a threshold 10^{-6}.

For comparison purpose, we first show the articles ordered by quality values (from high to low) derived from naive model in Table 2. Naive model measures quality by the number of words an article contains. “Germany” was given the highest rank as it contains 8912 words. It was followed by “People’s Republic of China” (7517 words). The smallest article “Anguilla” (821 words) was given the lowest rank. Since there are over 1000 contributors, we do not list them. It suffice to mention that the highest authority is given to “Tawkerbot2” who contributed 10,622 words. There were almost 600 contributors contributed not more than 10 words, and thus assigned with very small authority values (<0.00005).

The articles in decreasing quality values measured by the basic model and peer review model are shown in Tables 3 and 4 respectively. It is clear that the the rankings by quality are quite different among the three models. This can be attributed to the different ways quality is measured. Consider

Table 2. Articles in Decreasing Quality Values - Naive Model

Germany, People’s Republic of China, Hong Kong, Israel, Japan, Chile, Greece, Bosnia and Herzegovina, Afghanistan, Cyprus, Bangladesh, Bolivia, Argentina, Finland, Iran, Guyana, Belarus, Brazil, Egypt, India, Republic of Ireland, Indonesia, Eritrea, Hungary, Barbados, Iceland, Armenia, El Salvador, Austria, Abkhazia, Algeria, Italy, Bahrain, Haiti, Gibraltar, Colombia, Jamaica, Ecuador, Estonia, Isle of Man, Falkland Islands, Azerbaijan, Guernsey, Dominican Republic, Cameroon, Bermuda, Czech Republic, Botswana, Honduras, Angola, Cayman Islands, Guam, Equatorial Guinea, Chad, Guatemala, Benin, Albania, Faroe Islands, Grenada, Burkina Faso, Croatia, Burundi, Ghana, Greenland, American Samoa, Aruba, Belize, Brunei, Andorra, Dominica, Guinea-Bissau, French Polynesia, Comoros, Djibouti, Cook Islands, Gabon, Anguilla

Table 3. Articles in Decreasing Quality Values - Basic Model

Benin, Isle of Man, Cameroon, Grenada, Angola, Chile, Cook Islands, Cyprus, Equatorial Guinea, Barbados, French Polynesia, Burundi, Botswana, Cayman Islands, Estonia, Gabon, Falkland Islands, Guam, Bosnia and Herzegovina, Anguilla, Croatia, Azerbaijan, Bangladesh, Chad, Faroe Islands, American Samoa, Gibraltar, Bolivia, Belarus, Algeria, Haiti, Djibouti, Colombia, Abkhazia, Aruba, Jamaica, Comoros, Andorra, Dominican Republic, Greenland, Guinea-Bissau, Burkina Faso, Czech Republic, Republic of Ireland, Honduras, Dominica, Bermuda, Austria, Guyana, Guernsey, Iceland, Brunei, Belize, El Salvador, Albania, Hungary, Bahrain, Guatemala, Finland, People’s Republic of China, Israel, Afghanistan, India, Eritrea, Argentina, Japan, Armenia, Ghana, Indonesia, Germany, Ecuador, Hong Kong, Greece, Brazil, Iran, Egypt, Italy

Germany which receives the highest quality from the naive model. It is contributed by only 12 contributors, and most the contributed words comes from Buchanan-Hermit who contributed only the Germany article. Hence, Germany was ranked much lower by both the basic model (rank=70) and the peer review model (rank=62).

Benin’s quality was ranked top by both the basic model and the peer review model. It has 57 contributors, far more than Germany and some of them have fairly high authority values although it has only 1848 words, much less than the 8912 words in Germany. The peer review model ranked Benin at the top as most of its words have been seen by high authority contributors including Electionworld, Rick Block, and Tawkerbot2, each of which reviewed more than 1830 words of Benin.

Table 4. Articles in Decreasing Quality Values - Peer Review Model

Benin, Belarus, Bosnia and Herzegovina, Azerbaijan, Bangladesh, Cameroon, Djibouti, Chad, Burundi, Croatia, Falkland Islands, Grenada, Estonia, Cyprus, Cayman Islands, Dominican Republic, Faroe Islands, Iceland, Gabon, Isle of Man, Guatemala, Chile, Bolivia, Bermuda, Algeria, Gibraltar, Equatorial Guinea, Guernsey, People’s Republic of China, Indonesia, Afghanistan, French Polynesia, El Salvador, Guinea-Bissau, Armenia, Colombia, Aruba, Botswana, Belize, Honduras, Barbados, Finland, Czech Republic, Austria, Haiti, Greenland, Cook Islands, Bahrain, Andorra, Burkina Faso, Brunei, Jamaica, Guyana, Comoros, American Samoa, Republic of Ireland, India, Japan, Guam, Argentina, Angola, Germany, Anguilla, Dominica, Greece, Eritrea, Abkhazia, Israel, Ghana, Ecuador, Hong Kong, Albania, Brazil, Hungary, Iran, Italy, Egypt

5.3 Detailed Examples

In this section, we show some examples that illustrate the differences between the models. We first examine some articles that have major changes in their quality ranks as shown in Table 5. For each selected article, the table shows the top 5 contributors (in the case of basic model), and the top 5 reviewers/contributors (in the case of peer review model), as well as their authority information.

As shown in Table 5, the big improvement in rank from 56 to 1 for Benin was due to the consideration of authorities of contributors. In the case of basic model, the 1848 words in Benin's article were mainly contributed by very highly authoritative contributors (Kwamikagami, MJCdetroit, Gryffindor, Electionworld and YurikBot) who were ranked 1 to 29 by the basic model. In the case of peer review model, most of the 1848 words were seen (or reviewed) by highly authoritative contributors (Khoikhoi, Electionworld, Rick Block, Tawkerbots, and Gryffindor) who were ranked between 1 to 15 by the peer review model. On the other hand, Hong Kong which has a long article (6269 words, 3rd highest quality by naive model) was ranked very low by both the basic and peer review models. As shown in Table 5, the basic model ranked it very low due to less authoritative contributors who were ranked between 104 to 1045 by the basic model. The peer review model ranked Hong Kong very low because many words of Hong Kong were reviewed by non-authoritative contributors (ranked between 49 to 1061). The above two cases show that by considering the authority of contributors, the basic and peer review models are able to judge articles beyond looking at article size which may be deceptive.

Next, we show how basic model and peer review model can differ in their quality measurement. Angola was ranked 5th by the basic model but 61st by the peer review model. The disparity is due to the fact that Angola was contributed by a few very authoritative contributors including Rick Block, Electionworld, and Tawkerbot2, but most of the 1980 words were reviewed by non-authoritative contributors. People's Republic of China (7157 words) was ranked 62nd by the basic model and 29th by the peer review model. The basic model ranked its quality low because most of the words were contributed by non-authoritative contributors. The rank given by peer review model was higher due to authoritative reviewers such as Polaron (rank=5), Rick Block (rank=2) and Electionworld (rank=1) even though these contributors did not contributed much to the article.

We now consider the examples of contributors having major changes to their authority values as shown in Table 6. Contributor 172.183.219.102 contributed 898 words to one single article, Republic of Ireland. This is a fairly significant amount of contribution given that each contributor wrote on average 198 words (see Table 1). Hence, he was ranked 52nd by the naive model. Nevertheless, the same contributor was ranked much lower by the basic model and peer review model because he only contributed to one article which received below average quality ranks.

Another contributor GrinBot, on the other hand, had significantly higher authority ranks in both basic model (rank=15) and peer review model (rank=23). Although GrinBot did not contributed many words in total, some of these words went to quite highly ranked articles such as Cook Islands (rank=7) and French Polynesia (rank=11). GrinBot also reviewed some good quality articles such as Chad (rank=8) and Djibouti (rank=7). In the case of contributor Obli, the poor authority rank from the basic model could be attributed to Obli's only contributed article Indonesia receiving a poor quality rank (69).

5.4 Discussions

With the results obtained in our experiments, we can see that the basic model and peer review model are able to derive article qualities (and contributor authorities) from the collaboration information and edit histories through a mutual reinforcement approach. Compared to the naive model, the two proposed models are able to justify their quality and authority rank values better given that collaborative editing usually helps to improve article qualities. As the peer review model considers additional review component in collaborative editing, it can give high quality ranks to articles written by non-authoritative contributors but reviewed by highly authoritative contributors. This distinctive feature had clearly shown up in our experiments making it produced different results compared to the basic model. Depending on the nature of dataset and the importance of factors (i.e., authorities of authors, authorities of reviewers) to be used in quality measurement, one can choose between basic model and peer review model.

6 Conclusions

In this paper, we address the problem of measuring qualities of collaboratively created articles. Instead of depending on the article length, we propose two mutually reinforcement models known as the basic model and peer review model. The former directly considers the authorities of contributors and the amount of contribution by each contributor. The latter exploits the collaborative reviews of articles and confers qualities to articles based on the authorities of their reviewers. We have applied the two proposed models on a subset of Wikipedia dataset and compared their efficacies.

Following this research, we can identify a few future research directions for study. Firstly, there can be several variants of our proposed models by selecting different combinations of aggregation functions. The properties of these

	Quality Rank			Top 5 contributors/ # of contributed words/ Authority Ranks (Basic)	Top 5 reviewers or contributors/ # of reviewed words/ Authority Ranks (Peer Review)
	Naive	Basic	Peer Review		
Benin	56	1	1	Kwamikagami/896/29 MJCdetroit/194/7 Gryffindor/184/9 Electionworld/61/1 YurikBot/60/5	Khoikhoi/1845/16 Electionworld/1844/1 Rick Block/1843/2 Tawkerbot2/1836/4 Gryffindor/1717/9
Hong Kong	3	72	71	David.Mestel/6168/104 AntiVandalBot/28/263 Paul Christensen/26/1008 67.140.224.164/23/1010 BlueValour/7/1045	LakeHMM/6269/1061 Skinnyweed/6266/49 AntiVandalBot/6264/173 Paul Christensen/6236/780 71.224.192.29/6210/1045
Angola	50	5	61	Rick Block/766/2 Electionworld/538/1 72.57.64.90/547/71 203.131.142.189/35/50 Tawkerbot2/31/4	70.23.226.86/1979/1050 Rodri316/1978/814 205.250.15.1.220/1975/876 72.57.64.90/1973/339 203.131.142.189/1426/78
People's Republic of China	2	62	29	Ryz05/4646/390 Heqs/1891/425 Nlu/364/502 Tawkerbot2/154/4 Gaius Cornelius/92/33	141.153.114.88/7157/629 Tapion/7152/963 Polaron/7151/5 Rick Block/5165/2 Electionworld/5165/1

Table 5. Comparison Examples (Quality)

	Authority Rank			Top 5 contributed articles/ # of contributed words/ Quality Ranks (Basic)	Top 5 reviewed articles/ # of reviewed words/ Quality Ranks (Peer Review)
	Naive	Basic	Peer Review		
172.183.219.102	52	260	263	Republic of Ireland/898/44	Republic of Ireland/1219/56
GrinBot	337	15	23	Cook Islands/22/7 Faroe Islands/16/25 Chad/3/24 French Polynesia/3/11 Djibouti/2/32	Chad/1047/8 Faroe Islands/678/17 American Samoa/227/55 Djibouti/180/7 Aruba/106/37
Obli	114	826	191	Indonesia/421/69	Indonesia/2151/30

Table 6. Comparison Examples (Authority)

variants and their implications to the convergence of iteration computation can be studied in detail. Secondly, we plan to expand the experimental work on a larger dataset, such as the entire Wikipedia collection. This will allow us to study the scalability of the model in computation overheads. Finally, user evaluation of the computed quality and authority values can be carried out to further validate their effectiveness.

References

[1] Brian Amento, Loren Terveen, and Will Hill, *Does authority mean quality? predicting expert quality ratings of Web documents*, ACM SIGIR Conference on Research and Development in Information Retrieval, 2000, pp. 296–303.

[2] Francesco Bellomi and Roberto Bonato, *Network Analysis for Wikipedia*, Wikimania 2005 - The First International Wikimedia Conference, August 2005.

[3] Sergey Brin and Lawrence Page, *The anatomy of a large-scale hypertextual Web search engine*, International Conference on World Wide Web, April 1998.

[4] Jim Giles, *Internet encyclopaedias go head to head*, Nature (2005), 900–901.

[5] Zoltan Gyongyi, Hector Garcia-Molina, and Jan Pedersen, *Combating Web Spam with TrustRank*, Very Large Data Base Conference, September 2004.

[6] J. M. Kleinberg, *Authoritative sources in a hyperlinked environment*, Journal of the ACM **46** (1999), no. 5, 604–632.

[7] Marc Najork and Janet L. Wiener, *Breadth-first crawling yields high-quality pages*, International Conference on World Wide Web, 2001, pp. 114–118.

[8] L. Page, S. Brin, R. Motwani, and T. Winograd, *The pagerank citation ranking: Bringing order to the web*, Stanford Digital Library Technologies Project, 1998.

[9] Jakob Voss, *Measuring Wikipedia*, International Confererence of the International Society for Scientometrics and Informetrics, July 2005.

[10] Wikipedia, *Wiki — Wikipedia, The Free Encyclopedia*, 2006, [Online; accessed 4-July-2006].